JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1971, 4, 257-264 NUMBER 4 (WINTER 1971)

A COMPARISON OF STUDENTS STUDYING-BEHAVIOR PRODUCED BY DAILY, WEEKLY, AND THREE-WEEK TESTING SCHEDULES[1]

V. T. MAWHINNEY, D. E. BOSTOW, D. R. LAWS, G. J. BLUMENFELD, AND B. L. HOPKINS

SOUTHERN ILLINOIS UNIVERSITY

Time spent in contact with academic course materials as a function of two testing schedules was measured using college undergraduates in an introductory educational psychology course. A within-subject (ABAB) design was employed with all subjects to allow for both individual and group analyses. A study room equipped with an adjacent observation room enabled visual and auditory monitoring of student study behavior. Academic materials were exclusively available to students in the study room and records of durations and distribution of student study time were made by an observer behind a one-way mirror. Daily testing produced consistent duration of study behavior with regular attendance at the study room; weekly and three-week testing produced sporadic bursts of study behavior and frequent instances of non-attendance. The amount of study behavior occurring in weekly and three-week testing conditions increased as the test time drew near (scalloping). Results suggest that daily testing supports more consistent study patterns than do the two larger intertest intervals investigated.

Recently, complex experimental teaching programs similar to those developed by Keller (1968) and used by Ferster (1968), were found to produce final examination scores that were superior to scores occurring under traditional lecture methods (McMichael and Corey, 1969; Sheppard and MacDermot, 1970). Ferster (1968) presented cumulative records of oral examinations taken by students who were allowed to work on course materials at their own pace. Cumulative records of four subjects showed scalloped response patterns in which interviews became more frequent as the end of the course drew near.

Educational programs most often use one, or a combination of two general types of testing schedule. The first involves tests that occur at predictable, fixed, or variable intervals. These testing schedules may involve large intertest intervals such as mid-term and final examinations, or schedules with much shorter intervals such as daily quizzes. The second form of testing uses variable intertest intervals that occur according to a non-predictable schedule. This testing schedule is commonly called "pop quizzing".

The continued use of these testing schedules is probably based upon the personal preferences and casual observations of individual instructors. Numerous investigations have analyzed study methods using student questionnaires and self-reports. These studies, however, did not use direct methods of observation and typically did not investigate the amount or distribution of study time as a function of schedules of testing. Class (1935) found that students reported studying greater amounts of time in preparation for essay tests than they did in preparation for objective tests. Jones and Ruch (1928) used student self-reports and found that students who scored low on intelligence tests reported that they studied an average of five hours a week more than students who scored high on intelligence tests. Hutchinson and Conrad (1926) investigated the amount and distribution of time spent studying with students who recorded their own study times. Time spent engaging in extracurricular activities and study behavior were reported to increase together as the end of the semester approached.

[1]The authors wish to express their gratitude to members of the Guidance and Educational Psychology Department of Southern Illinois University who willingly cooperated in the conduct of this research. We are particularly indebted to Drs. John Cody, John Mouw, and Keith McNeal for their efforts in providing research space used in this study. We also wish to thank Dr. Beth Sulzer for her comments and suggestions on research design and manuscript preparation. Reprints may be obtained from V. Thomas Mawhinney, Department of Psychology, Indiana University, South Bend, Indiana, 46615.

The present study investigated the relationship between three testing schedules and the distribution of study behavior directly observed during the intertest interval. Testing schedules with one week and three-week intertest intervals were contrasted with daily testing in order to see if scallop effects similar to those observed in the records of cumulative interviews (Ferster, 1968) would also occur in directly observed study behavior.

EXPERIMENT I

This involved the direct observation of study behavior and compared the amount and distribution of time spent studying as a function of daily and weekly testing schedules.

METHOD

Subjects

Six females, 18 to 51 yr old and two males 21 and 23 yr old, had passed a prerequisite course in general psychology and were undergraduate or graduate students at Southern Illinois University. All subjects were members of a large introductory lecture course in educational psychology who volunteered to participate in a special readings section for course credit. Participation in the special readings section satisfied all course requirements for the regular lecture section. The criteria for selection of subjects were: (1) that they had no previous experience in behavioral psychology, (2) that they had at least an overall C average, and (3) that they had no academic or work commitment during the hours the study room was available for their use.

Materials

The study materials were composed of the entire text of C. B. Ferster and M. C. Perrott's *Behavior Principles* (1968) and selected chapters from B. F. Skinner's *The Technology of Teaching* (1968). Identifying titles and names were removed from the text materials. Each subject was assigned a specific set of materials for which he was responsible for the duration of the study.

An attempt was made to equate the length of daily reading assignments insofar as the organization of textual material allowed. This was done before the study began. Daily reading assignments averaged approximately 15 pages. The *Behavior Principles* text is partially programmed with each major subject matter section followed by a set of behavioral objectives related to that section. The objectives defined for the students some of the more important skills that they were to acquire from the readings. In the case of the *Technology of Teaching,* probes or behavioral objectives were prepared and inserted into the text materials. Ferster and Perrott's *Behavior Principles* was presented first, followed by Skinner's *Technology of Teaching.*

Study Area

A well lighted, air-conditioned 17 x 25 ft seminar room served as the study area for the subjects. Subjects sat in desk chairs or around a large table at one end of which the numbered study materials had been placed. Notices forbidding admittance to everyone other than the subjects during the designated session hours were posted at all entrances to the room.

Response Definition

Time spent "studying" was defined as that period from when a subject seated himself at the table with the text material in front of him until he got up from the chair. The observer recorded the total number of minutes each student was seated in front of the study materials during each study session. Behaviors such as glancing about the room, counting of pages, lighting a cigarette, *etc.* were scored as being part of the total complex of study behavior and were included as "time spent studying." When a subject failed to attend the study room, zero minutes of study time was recorded for that session. Observers scored start and stop times by rounding to the nearest half minute.

Observers and Observation Area

The observers were four males, two graduate students and two faculty members of Southern Illlinois University. Observations were made by a single observer during each three-hour study session. Adjacent to the study room was an 8 x 11 ft room containing a 35 x 43 in. one-way glass window that afforded an unobstructed view of the study area. The observer sat at a table behind this window and recorded the times that students began and stopped studying. A microphone in the wall of the study area permitted continuous monitoring of the study room noise level. Two

observers sat in the darkened observation room and independently recorded student study times from the same clock. Observers were seated in a manner that reduced the likelihood of copying. Per cent agreement scores were calculated by dividing the number of minutes in the larger observation into the number of minutes in the smaller observation for each subject when inter-rate differences occurred. An inter-observer reliability check was taken during Session 11 for all subjects and yielded per cent agreement scores ranging from 92 to 100.

Procedure

Instructions to subjects. All subjects received a course outline that described the procedures to be followed in the course, such as use of the study area, the relationship of behavorial objectives to the text material, and the relationship of accumulated quiz points to the final grade.

Letter grades were based upon accumulated quiz points earned in the course. Students could accumulate a maximum of 180 points in nine weeks. Grades of A were assigned to 160-180, B to 140-159, C to 120-139, and D to 100-119 total points.

The study area was made available to the subjects from 3 p.m. to 6 p.m. on Monday through Thursday. Subjects were permitted to come or go from this area at any time during this period, or not to attend at all. In addition, the following requests were made and enforced: (1) verbalizations were to be kept to a minimum, (2) subjects were to avoid bringing visitors into the study room, (3) subjects were to study only the course materials provided in the room, and (4) materials or notes pertaining to the materials could not be taken from the study room.

Testing. Daily tests were worth five points each and weekly tests, which covered four daily assignments, were worth 20 points. Test were either multiple-choice, true-false, or fill-in-the-blank; each question was normally worth one point. Test questions were constructed as the experiment progressed using the behavioral objectives incorporated with each reading assignment. Each test period was followed by a voluntary discussion period in which test items and study materials were reviewed and discussed. Subjects were tested over the study materials on Tuesday through Friday if the testing schedule was daily, or on Friday only if the schedule was weekly.

A within-subject (ABAB) research design was employed with all subjects. Testing conditions were alternated in the following manner. During the first three weeks, appropriate daily reading assignments were made available in the study room for one day at a time. On the following day, a five-point quiz was given over those materials. During the fourth and fifth weeks, all reading materials for the three weeks were made available in each study session. A 20-point test was given on Friday of each of these weeks. During the sixth week, daily quizzes were again given but all study materials for the week's quizzes were made available on Monday. The length of each reading assignment was indicated by notes in the text material. In weeks seven and eight, the weekly testing schedule was again employed. In the ninth week, conditions were again changed to daily testing with all materials for the week available beginning on Monday.

RESULTS

Variability in day-to-day minutes studied was observed in each subject's records. Overall amounts of time studied varied among students. Subject #2 studied the least with a mean duration of 31 min per day daily study times ranging from 17 to 68 minutes. Subject #8 studied the most with daily study times ranging between a high of 140 and a low of 59, with a mean of 108 min. Beginning with Session 12, 20-point tests were given after four daily study sessions (weekly testing). The change in conditions from daily to weekly tests produced a marked increase in the variability in minutes studied from session to session. This variability is most notable in the high frequency of absences that occur early in the weeks when tests are given on the last day of the week. During Session 12, the first study session of the weekly testing condition, three of the eight subjects failed to attend the study room. Three subjects failed to attend Session 13, one subject Session 14, while all subjects studied in Session 15, the last opportunity before the test. The performances of Subjects #6 and #7 most dramatically changed with the shift from daily to weekly testing conditions. The data plots for these subjects include scallops, with the minutes studied per

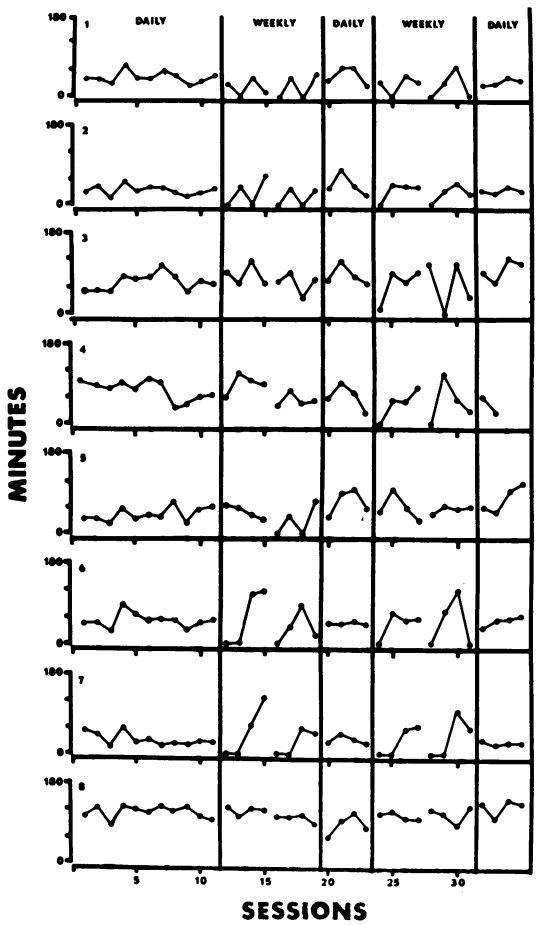

Fig. 1. Minutes studied per session for each subject during alternating conditions of daily and weekly testing in Experiment I.

study session increasing with increasing temporal proximity to the weekly test. There were no systematic changes in the study patterns of Subject 8 throughout the experiment.

A return to the daily testing condition in Session 20 produced daily study patterns for most subjects, typical of those occurring during Sessions 1 to 11. Again, no absences from the study room occurred. When the weekly testing situation was repeated during Sessions 24 to 31, absences again occurred more frequently early in each week and decreased with increasing proximity to the test. The average number of minutes studied per session for all subjects during daily testing was 56.28 min in Sessions 1 through 11, 58.28 min in Sessions 20 through 23, and 63.50 min in Sessions 32 through 35. Average amounts of the time studied were slightly lower during weekly testing with 52.89 min in Sessions 12 through 19 and 50.89 in Sessions 24 through 31.

Discussion

The results include distinct differences in the distribution of studying produced by daily and weekly testing conditions. Every subject attended all of the study sessions during daily testing conditions. Weekly testing produced more variability in both amounts of daily study time and attendance in the study room in all but one subject. Two subjects tended to study more nearer the end of the intertest interval. The average total amount of time studied was slightly greater during daily testing conditions than during weekly testing.

EXPERIMENT II

A second experiment employed three-week testing schedules to investigate the effects of longer intertest intervals on students' study behavior. A direct replication of the study behavior produced by daily quizzing in Experiment I was contrasted with study behavior that occurred during three-week testing intervals. The three-week interval was considered to approximate frequently employed classroom testing schedules and allowed for an ABAB repeating conditions research design within a 10-week academic quarter.

Method

Subjects and Materials

Six males and six females, ages 19 through 26 yr, volunteered from an undergraduate course in educational psychology. Criteria for selection of subjects, reading materials, study and observation areas, instructions to the subjects, and response definitions were identical to those employed in Experiment I. Except for occasional minor revisions on specific test questions, all items were constructed before the beginning of the experiment. Inter-observer reliability scores were computed in the same manner as in Experiment I. Reliability checks were taken in Sessions 16, 28, and 40. Per cent agreement scores ranged from 96 to 100.

Procedure

Reading materials were introduced such that at any time during the course at least two subsequent weeks of reading materials were available to the students. The accumulated

reading materials remained available throughout the duration of the course. Two testing schedules were presented alternately, daily testing and testing after a three-week interval. During the first two weeks, subjects were tested daily with five-point quizzes. Beginning with the third week, a three-week testing schedule was employed. At the end of this three-week interval, a 60-point quiz was given. All reading materials for the three-week testing interval were available on the first day of the third week. During the sixth and seventh weeks the daily testing schedule was again in effect. After the seventh week, the three-week schedule was again used for the last three weeks of the course.

Results

Figure 2 shows individual study time in minutes per session for Subjects 9 through 20. Under daily testing conditions in Sessions 1 through 8, the number of minutes studied per session for each student indicated consistent daily attendance in the study room similar to that observed in Experiment I. The change in testing conditions from daily tests to tests every three weeks beginning with Session 9 produced a marked increase in the variability in minutes studied from session to session for every individual. When the testing schedule was changed from a daily to a three-week schedule, every student failed to come to the study room on at least some days. Nearly all subjects produced scalloped response patterns in either one or both of the three-week test intervals.

In Session 21, the first session following the return to daily testing conditions, a double reading assignment was accidently given that correlated with the unusually large number of minutes studied on that day for Subjects 14, 15, 17, and 18. During Sessions 21 through 28, study patterns were similar to those observed in the first daily testing condition. The return to a three-week testing schedule, beginning with Session 29, produced scallop patterns of studying similar to those observed in the first three-week testing condition.

Each data point in Figure 3 represents the mean number of minutes studied per session for all subjects during each of the testing conditions. Means were calculated by summing minutes studied per session for all subjects and dividing these totals by 12. In contrast to the relatively stable performance during daily testing conditions, the group average number of minutes studied increased with increasing temporal proximity to the three-week test. During Sessions 1 through 8, the mean number of minutes studied was 64. In Session 9, the average minutes studied decreased to approximately 20 min, then subsequently increased to approximately 115 min in Session 20, the last session preceding the three-week

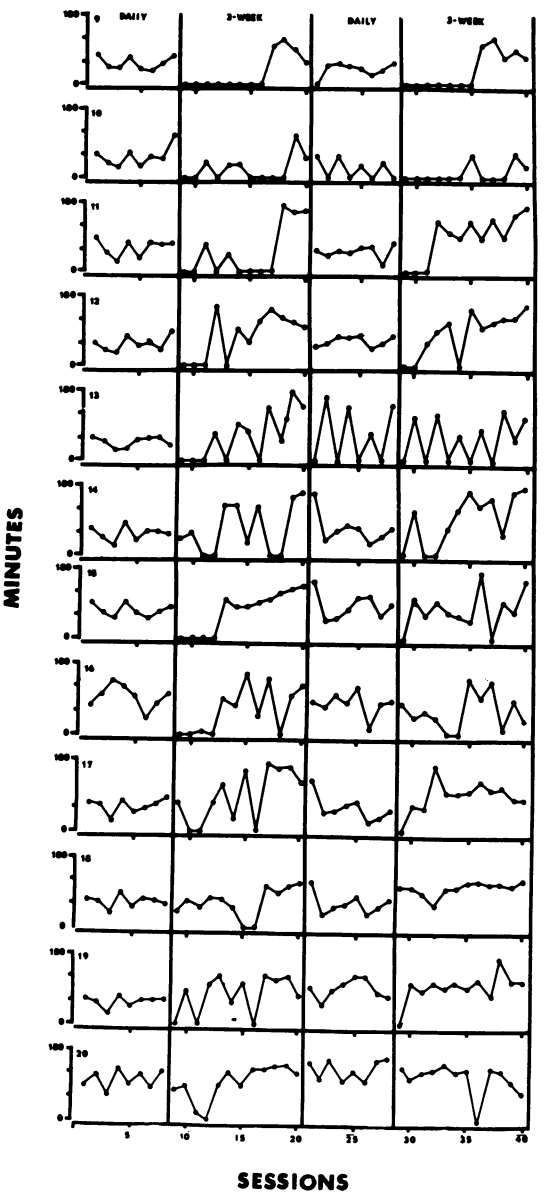

Fig. 2. Minutes studied per session for individual subjects during alternating conditions of daily and three-week testing in Experiment II.

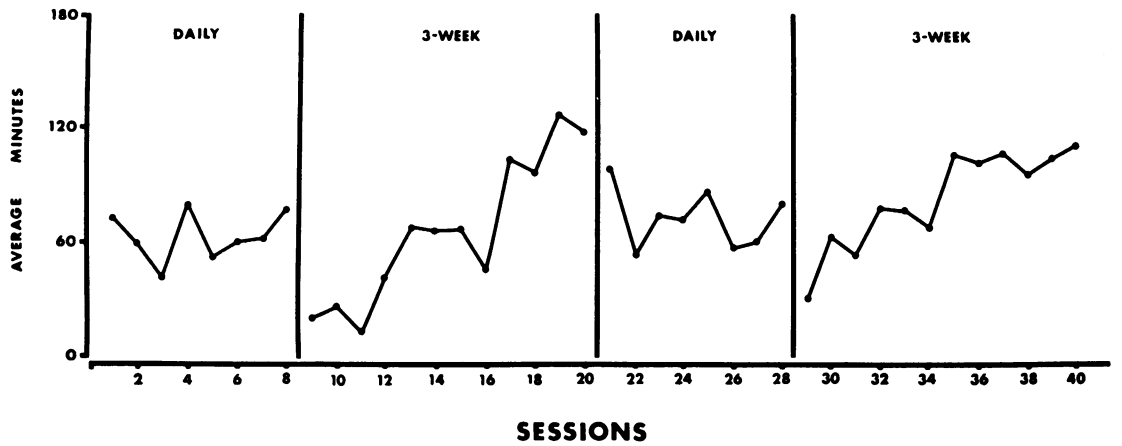

Fig. 3. Average minutes studied per session for all subjects during alternating daily and three-week testing conditions in Experiment II.

test. A subsequent return to daily and three-week conditions again revealed study patterns characteristic of those observed in the first two testing conditions, except that the rate of increase in mean number of minutes studied per day was lower in the final three-week testing condition. The average number of minutes studied per session for all subjects during the two daily testing conditions was 64.07 min in Session 1 through 8 and 73.33 min in Sessions 21 through 28. The first three-week testing condition produced an average of 63.49 min per session during Sessions 9 through 20, with the highest average of 83.55 min occurring in the last three-week testing interval, Sessions 29 through 40.

Discussion

The consistent daily study patterns produced by daily testing in Experiment I were replicated in Experiment II. Nearly all subjects in Experiment II produced clear scalloped study patterns under three-week testing conditions. Subjects 10 and 13 exhibited unusual study patterns after Session 21, and later admitted that they had made commitments that conflicted with daily study sessions, thereby violating course rules.

GENERAL DISCUSSION

Experiments I and II demonstrate a functional relationship between two testing schedules and the resulting distribution of study behavior. In Experiment I, time studied per session was relatively inconsistent during weekly testing. This condition also produced absenteeism. The number of minutes studied by several students increased as the time of testing drew near. Experiment II was conducted to determine if the scallop effects observed in Experiment I would become more pronounced at larger intertest intervals. The three-week testing interval employed in Experiment II also produced a great deal of variability in time spent studying per day by each student and frequent days of no studying. In addition, pronounced scallops were observed to occur in the study behavior of most students during the three-week testing interval. Due to constant change in the amount and difficulty of reading materials, between condition comparisons of the amount of time studied is difficult. In Experiment I, the average number of minutes studied per session by all subjects during each of the daily testing conditions was slightly higher than those averages occurring under the two weekly testing intervals. This relationship was not as consistent in Experiment II, when the highest average occurred in the last three-week testing condition. When Subjects 10 and 13 were questioned regarding the unusual number of days in which they did not study after Session 21, it was found that they had made commitments, in violation of course rules, that conflicted with attending the study sessions.

An initial attempt was made to equate daily reading assignments with respect to length and level of difficulty. The choice of daily assignment length was largely defined, however, by the subject matter divisions in both *Be-*

havior Principles and *The Technology of Teaching*. The *Behavior Principles* text is arranged into short readings, followed by sample questions over important aspects of each reading. Generally, quiz questions in the study were taken from these sample questions. A page in the daily readings was composed primarily of either, or a combination of prose, content outlines, or graphs. As many pages were combinations of prose, graphical representations, and open space, equation of assignments based on the number of pages alone was deemed inadequate. Therefore, in an effort to maintain continuity in single assignments the readings followed by each set of objectives were used as daily assignments. These assignments averaged approximately 15 text pages. This average number of pages was also used in dividing the *Technology of Teaching* as well as the remaining selected readings into daily assignments. Two factors possibly produced day-to-day variability, the differences in the length of readings and the constantly changing subject matter made comparison between specific days, as well as between conditions, difficult. Between-subject comparison with respect to changes in the absolute amount of reading on successive days in the daily testing condition suggested a relationship between number of pages assigned and the length of time the students spent reading the materials. In Experiment II, a double reading was assigned accidently for Session 21. The high number of minutes spent studying by Subjects 14, 15, 17, and 18 in this session suggests a relationship between length of reading assignment and the amount of studying time.

Two considerations contributed to the absence of analysis of test performance in the present study. Each daily quiz was composed of no more than five objective questions. Thus, if a subject missed one question, he missed 20% of the test items. A score of four on one day could not be easily compared to a score of four on a different day as the complexity or difficulty of content material was not easily quantified and therefore varied in undetermined ways throughout the study.

Scalloped study patterns occurring in three-week testing conditions approximated those cumulative interview records presented by Ferster (1968). Recently, similar scalloped cumulative records were presented for student acquisition of points in a programmed college course (Lloyd and Knutzen, 1969). Although these studies did not directly measure study behavior, the patterns of cumulative interviews (Ferster, 1968) and cumulative points earned (Lloyd and Knutzen, 1969), were probably a function of scalloped study behavior similar to that observed in the one-week and three-week testing schedules in the present study.

One weakness of the "self-paced" course is that student response rates occurring in the early portion of the course are frequently minimal. Lloyd and Knutzen observed a direct relationship between the time at which students began to respond and their final grades achieved in the course. Those students who began work later in the course, accomplished fewer tasks and received lower grades. Malott and Sviniki (1968) reported that daily assignments and frequent deadlines seemed to produce steady rates of reading behavior during an introductory psychology course. Daily deadlines for study behavior imposed by the daily testing schedule in both of the present experiments produced consistent rates of study behavior and lends support to Malott's and Sviniki's suggestion.

Large intertest intervals, such as midterms and finals, are often used in the classroom. The use of these schedules of testing is likely a result of instructor convenience or the assumption that this form of testing is sufficient to motivate the student to study in a consistent fashion for the duration of the course. The present results seriously question the effectiveness of large-interval testing schedules as being adequate devices for producing consistent student study behavior.

REFERENCES

Class, E. C. The effect of test announcement on students preparation. *Journal of Educational Research,* 1935, *XXVIII,* 358-361.

Hutchinson, R. and Conrad, M. What's in a college week? *School and Society,* 1926, **24,** 769-772.

Jones, L. and Ruch, G. M. Achievement as affected by amount of time spent in study. 27th Yearbook. *National Society for the Study of Education,* Pt. II, 1928, 131-134.

Keller, F. S. Good-bye teacher. *Journal of Applied Behavior Analysis,* 1968, **1,** 79-89.

Lloyd, K. E. and Knutzen, N. J. A self-paced programmed undergraduate course in the experimental analysis of behavior. *Journal of Applied Behavior Analysis,* 1969, **2,** 125-133.

Malott, R. W. and Svinicki, J. G. Contingency man-

agement in an introductory psychology course for one thousand students. *The Psychological Record,* 1969, **19,** 545-556.

McMichael, J. S. and Corey, J. R. Contingency management in an introductory psychology course produces better learning. *Journal of Applied Behavior Analysis,* 1969, **2,** 79-83.

Sheppard, W. C. and MacDermot, H. C. Design and evaluation of a programmed course in introductory psychology. *Journal of Applied Behavior Analysis,* 1970, **3,** 5-11.

Received 5 October 1970.
(Revised 30 June 1971.)